DOCUMENT RESUME

ED 457 212 TM 033 310

AUTHOR	Kulinna, Pamela Hodges; Cothran, Donetta; Regualos, Rey
TITLE	Development of an Instrument To Measure Students' Views of Classroom Management.
PUB DATE	2001-04-00
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Behavior Patterns; *Classroom Techniques; Content Validity; *High School Students; High Schools; Higher Education; Learning; *Middle School Students; Middle Schools; Physical Education; Reliability; Scores; *Student Attitudes; Test Construction; *Undergraduate Students
IDENTIFIERS	*Experts

ABSTRACT

The purpose of this study was to develop an instrument to examine students' behaviors in physical education classes that might have an impact on classroom management and student learning. The study was conducted in multiple phases, including instrument development, preliminary study, content validity study, and a reliability and validity study. Participants for the content validity study were 27 experts in sport pedagogy. They were sent items from the management instrument and were asked to organize them into three severity categories. Participants for the reliability and validity study were 2,309 middle and high school students from 2 states. Statistical results indicate that an instrument that produces reliable and valid scores was developed to measure students' perceptions of classroom management in physical education settings. Investigating students' views in this area will lead to a better understanding of management issues and behaviors that interfere with students' ability to learn. (Contains 3 tables and 38 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 212

DEVELOPMENT OF AN INSTRUMENT TO MEASURE STUDENTS' VIEWS OF CLASSROOM MANAGEMENT

Pamela Hodges Kulinna, Wayne State University

Donetta Cothran, Indiana University

Rey Regualos, Wayne State University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Hodges

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Send correspondence to:

Pamela Hodges Kulinna
Assistant Professor
Wayne State University
College of Education
263 Matthaei Building
Detroit, MI 48202
Phone: (313) 577-9847
Fax: (313) 577-9300
E-mail: P.Kulinna@wayne.edu

Running Head: Classroom Management Instrument

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033310

Paper presented at the annual meeting of the American Educational Research Association, Seattle, April 2001.

BEST COPY AVAILABLE

Abstract

The purpose of this study was to develop an instrument to examine students' behaviors in physical education classes that might impact classroom management and ultimately student learning. The study was conducted in multiple phases, including instrument development, preliminary study, content validity study, and a reliability and validity study. Participants for the content validity study were 27 experts in sport pedagogy. They were sent items from the management instrument and asked to organize them into three severity categories. Participants for the reliability and validity study were 2,309 middle and high school students from two states. Statistical results indicate that an instrument that produces reliable and valid scores was developed to measure students' perceptions of classroom management in physical education settings. Investigating students' views in this area will lead to a better understanding of management issues and behaviors that interfere with students' ability to learn.

DEVELOPMENT OF AN INSTRUMENT TO MEASURE STUDENTS' VIEWS OF CLASSROOM MANAGEMENT

Classroom management problems have inundated teachers from time immemorial and will likely continue to be an important factor in the future. Teachers of all experience levels report that classroom management is one of the most challenging and disturbing aspects of teaching (McCormack, 1997) and in public discourse of issues of education, school discipline has emanated as the school area of most concern (Duke & Jones, 1984). Management's central location in educational and public discussion is well deserved as management and organization are at the very heart of teacher effectiveness (Brophy, 1983).

Classroom management refers to a broad range of teacher behaviors and one of the most critical is managing student behavior. The consequences of student misbehavior are multifaceted (Vogler & Bishop, 1990). Student troublesome behavior interferes with student academic learning time which is highly correlated with student achievement (Doyle, 1986), can make students with good behavior feel uncomfortable when misbehaving students are admonished (Kounin, 1970) and may lead to teacher dissatisfaction, stress and burnout (Depaepe, French, & Lavay, 1985; Wolfgang & Glickman, 1980). Adequate classroom management is a prerequisite for protecting students' physical, psychological, and social well being, as well as to achieving instructional objectives (Clarizio, 1976).

Due to its importance, student misbehavior has been the focus of much classroom research. The most common approach to investigating student misbehavior has been to survey teachers' perceptions of disruptive behaviors occurring in classes. Teachers' perceptions of the most frequent and troublesome behaviors in secondary (Houghton, Wheldall, & Merrett, 1988; McNamara, 1985; Ziv, 1970) and primary school settings (Fields, 1986; Wheldall & Merrett, 1988) have been investigated. Findings from a number of studies of British teachers at both the elementary and secondary level suggest that over half of teachers studied spent more time on order and control than they thought they should. The most troublesome behaviors were those that were reported the most frequently with two problems standing out--talking out of turn and hindering others--at both levels. From these findings, it was concluded that the majority of the classroom behavior problems experienced by elementary and secondary teachers were similar and although teachers were bothered by the behavior of their students, the most common and troublesome behaviors were relatively trivial (Houghton, et al., 1988; McNamara, 1985; Wheldall & Merrett, 1988). In related work, Borg and Falzon (1989a, 1989b, 1993) also showed the important role of context variables, in particular, teacher and pupil characteristics (especially gender) in influencing teachers' perceptions of problematic behaviors.

Student misbehavior also has been assessed through classroom observation and interviews. Observation techniques involve recording actual student misbehaviors during classes (e.g., Kennedy, 1982; Wragg, 1995). While this method may be the one of the

most accurate, it is exceedingly time intensive and difficult to administer across teachers and a wide range of school environments. Interview data collection suffers from similar challenges.

While survey, observational, and intervention studies focusing on teachers' perspectives and behaviors are important and have provided key insights into classrooms. It is unexpected that classroom management has rarely been investigated from the perspective of the student. Particularly since it is student behaviors that serve as the genesis of teacher action and reaction. It also is important to examine the student perspective to determine if teacher behaviors are effective. Students are actively involved interpreting and influencing the learning environment (Nicholls, 1992). Doyle's (1977) cognitive mediation paradigm recognizes that students' perceptions of and reactions to teachers' actions, rather that the teachers' actions alone, determine the level of student engagement and learning that occurs. Students and teachers may not assign the same meaning to events that occur in the classroom, due to the different experiences, expectations and needs those individuals bring to the classroom. There have been discrepancies observed among teachers' and students' perceptions of the same teaching-learning events in classrooms (Cullingford, 1991; Farrell, Peguero, Lindsey, & White, 1988; Wragg, 1995) and physical education settings (Cothran & Ennis, 1997; Solmon et al., 1998).

In one of the few management studies to address this potential mismatch between teachers' and students' management perspectives, Zeidner (1988) found students' and teachers' mean ratings of the perceived severity of various classroom management strategies were significantly different for half of the strategies, while they agreed on the relative severity of the most stringent forms of control. Similarly, Wragg (1995) found similarities and differences between teachers' and elementary students' interpretations and the meaning they assignment to classroom management events.

It seems likely that students and teachers might have different views of class management events in physical education settings, but there is a paucity of research in physical education that: (a) addresses troublesome behaviors occurring in classes and (b) involves the student perspective. To address the lack of research in this area, the purpose of the current study was to develop an instrument that produces reliable and valid scores to examine students' views of troublesome student behaviors in physical education classes impacting classroom management and ultimately student learning.

There have been a number of studies performed on the validity of data collection instruments in physical education (e.g., Keating, Kulinna, & Silverman, 1999; Kulinna & Silverman, 1999); however, no validation studies have been performed on instruments designed to measure students' views of troublesome behaviors in physical education classes. It is necessary to demonstrate that instruments produce reliable and valid scores in order to accurately measure psychological and social characteristics. Good instrumentation is needed in order for studies to produce useful information about the questions of interest (McDonald, 1999). Further validation studies are needed on

instruments used in educational research (Silverman & Subramaniam, 1999; Stewart & Destache, 1992).

An instrument that produces reliable and valid scores to assess students' views on troublesome behaviors occurring in physical education classes will help teachers and administrators to begin to understand the troublesome behaviors that take place in classes from the student perspective. Understanding troublesome student behaviors from this perspective will provide teachers with valuable information that they can use to change their behavior management strategies. Once the troublesome behaviors have been identified, strategies also can be developed and tested to reduce those undesirable behaviors in classes.

Method

This study was conducted in multiple phases. The first stage involved the development of a classroom management instrument to assess students' views of troublesome student behaviors in physical education classes. It also involved a series of pilot studies and instrument revisions. Content-related evidence for the management instrument was assessed by experts in physical education pedagogy in the second phase. Finally, middle and high school students were recruited to participate in the third phase, assessing the reliability and validity of scores produced by the instrument.

Phase One: Instrument Development

An instrument was developed to assess students' views of troublesome student behaviors that occur in physical education classes. The instrument was designed to measure the range and frequency of student misbehaviors present in classes. The items were compiled from a thorough review of the literature on classroom management. Items were gathered from studies addressing the range, frequency, and seriousness of behaviors in classroom settings (Borg, 1998; Houghton et al, 1988; Wheldall & Merrett, 1988) and physical education environments (Kennedy, 1982; McCormack, 1997). New items also were generated to represent unique behaviors to physical education programs (e.g., forgets gym clothes) inspired by the classroom management literature in physical education (Kennedy, 1982; McCormack, 1997; Vogler & Bishop, 1990). The instrument features a 5-point Likert-like scale response format (from never to always) for each of the items.

Phase One: Preliminary Studies

The instrument was then used in a series of pilot studies. There were a number of purposes driving the preliminary studies including: (a) to assess the reliability and validity of the scores produced by the instrument; (b) to add new items generated by students to represent behaviors occurring in physical education classes that were not included on the original instrument; and (c) to test the feasibility of the instrument. For each of the three pilot studies, students and parents provided informed consent and the instruments were administered during regular physical education classes.

Pilot study one: method and results. Participants for the first pilot study were 115 undergraduate students from two physical education teacher education programs in the midwest. There was adequate representation from both genders with 48 (42%) female and 67 (58%) male participants. The majority of the students reported their ethnic background as European-American (93 students--80.9%) or African-American (15 students--13.0%), with the remaining 6.1% reporting Hispanic-American, Asian-American, or multiracial/other. Due to the large number of items that were generated during instrument development, the 56 unique items were randomly separated into two versions of the classroom management instrument (28 items each) for this first pilot study. Students completed one of the two initial instruments, with 63 students completing form A and 52 students completing form B. It took students approximately 10 minutes to complete the survey. Data analysis included Cronbach's alpha coefficient as a measure of the internal reliability of the items on the instrument. Frequencies and descriptive statistics were also calculated for the instrument's items.

Mean scores on the items ranged from 1.31 (sd=.47) to 4.06 (sd=.75) on the one-to-five Likert-like scale with one representing never and five indicating always. Cronbach's alpha coefficients showed a high level of internal reliability among the items. The alpha coefficients for forms A and B were .91 and .95, respectively. Based on descriptive statistic results showing all behaviors occurring (to some degree) in classes, all of the items on both instruments were retained and combined into the second version of the instrument.

Pilot study two: method and results. Participants for the second pilot study were 180 undergraduate students from four physical education teacher education programs representing the midwest and eastern part of the country. Female (77) and male (97) participants were both well represented in the sample. Similar to the first pilot study, the majority of the students reported their ethnicity as European-American (148--85.1%) or African-American (15--8.6%), with the remaining 6.3% reporting Hispanic-American, Asian-American, or multiracial/other. In the second pilot study, the revised 56-item instrument was administered to students during classes in physical education departments. The participants also were asked to identify student misbehaviors in physical education classes that were not included on the instrument. Data analyses included an assessment of the internal reliability of the items on the instrument (i.e., Cronbach's alpha), frequencies and descriptive statistics.

Results from the second pilot study showed similar findings to the preliminary study. All of the troublesome behaviors were reportedly present in classes. Mean scores on the items ranged from 1.41 (sd=.69) to 4.02 (sd=.82). The combined instrument showed a high level of inter-item agreement with an alpha coefficient of .98. The undergraduate students identified three additional behaviors that were subsequently added to the instrument, including: (a) displays gang symbols, (b) threatens others, and (c) obscene gestures.

It took students approximately 15-20 minutes to complete the instrument. Teachers were left with adequate class time to meet other lesson objectives during the

session. Student responses demonstrated their understanding of the items on the instrument. Thus, it was hypothesized that it would be feasible to administer the instrument during physical education classes.

Pilot study three: method and results. In the final pilot study, 155 high school students (65 male and 90 female) from two schools (in different districts) completed the revised 59-item instrument. The majority of the students reported their ethnicity as European-American (92.9%). A small number of participants also reported their ethnicity as multiracial/other (3.9%), Hispanic-American (1.9%), or Asian-American (1.3%). Most of the participants were in the 9th grade (94.8%). The management instrument was administered at the schools by at least two members of the research team using a standardized administration protocol. The protocol was slightly modified after this pilot study with ninth grade students. Data analyses included frequencies, descriptive statistics, and calculating the Cronbach's alpha coefficient.

Results for third pilot involving high school students completing the revised 59-item instrument, included mean values for the items ranging from1.16 (sd = .53) for item 21 "bringing weapons to classes" to 4.57 (sd = .85) for item 10 "talking". The alpha coefficient indicated a high level of inter-item agreement with a value of .96. The instrument demonstrated that it could produce reliable scores in a population of high school students.

Results from the third pilot study also indicated that it was feasible to administer the management instrument in physical education classes. Students completed the instrument in 10-15 minutes with responses demonstrating their understanding of the items on the survey. Based on these results and the content-related evidence from the experts, the final version of the management instrument was created and used in the reliability and validity study.

Phase Two: Content-Related Evidence

Content-related evidence for the management instrument was assessed by experts in physical education pedagogy in the second phase. Thirty-nine experts in physical education pedagogy each with a terminal degree related to physical education pedagogy and experience with public school settings were contacted electronically regarding participating as experts for this study. Twenty-seven experts agreed to participate (69.2%) representing 17 states. Sixteen of the experts were females and 11 were males. Most of them reported their ethnicity as European American/Caucasian (88.9%). Experts reported years of experience in their physical education pedagogy positions as less than three years (16.7%), three-to-five years (5.6%), six-to-ten years (38.9%), eleven-to-twenty years (27.8%), or 21 or more years (11.1%). They were sent a list of the 59 student behaviors along with definitions for classifying the items into the categories of mild, moderate, and severe behaviors (see Table 1 for a listing of the definitions). The definitions were modified from Vogler and Bishop's (1990) classifications of mild, moderate and severe disturbances. Experts also provided demographic information and any comments that they had about the items and/or the instrument.

Expert instrument preliminary studies: method and results. Two small preliminary studies were conducted on the expert instrument to assess the classifications and organization of the instrument. Pilot study one included eight experts in physical education pedagogy. The pedagogy faculty completed the expert instrument, assessed the descriptions, and listed any questions or concerns they had regarding the items or management instrument. The percentage of agreement of the experts with the items was calculated for all of the 59 items.

Results from the first expert pilot study suggested that the definitions provided for mild, moderate and severe student disturbances were overly vague, only 43 of the 59 items (78%) showed at least 75% agreement among the experts. The classifications were elaborated upon for the second pilot study. Three of the management items also were modified, for example, item 32 was changed from "can't stand to lose/fail" to "gets upset when loses/fails". Five experts in physical education pedagogy participated in the second pilot study. Results showed much stronger agreement levels. The experts had $\geq$80% agreement on all of the items. Based on the results from the second pilot study, the final version of the expert instrument was created and used to assess content-related evidence with the physical education pedagogy experts.

Content-related evidence data analyses. Agreement on severity of behaviors was assessed by investigating the difference between mean rating of an item derived from the experts with the item's gold standard, i.e., mild, moderate or severe--set by the researchers based on Vogler and Bishop's (1990) definitions. An item showing a difference of at least .20 represents a significant discrepancy at or beyond the .05 level. The reliability of the experts' ratings was assessed by calculating an overall intraclass correlation coefficient.

Phase Three: Reliability and Validity Study

Participants were recruited by contacting 14 school districts, principals, and then the teachers at each school. When approval was gained at all levels, the teachers were sent the informed consent forms to distribute to students and their parents. A date also was scheduled for the research team to come to the school to administer the survey to the agreed upon classes using the standardized instrument administration protocol. The participants for the final phase were 2309 middle and high school students from 18 schools in two midwestern states. Students and parents provided informed consent and the management instrument was completed during physical education classes. There was similar representation from both genders (m=1145, f=1131). Students had various ethnic backgrounds including European American/Caucasian (n=1300), African-American (n=744), Hispanic (n=84), Chaldean (n=29), Asian-American (n=16), and multiracial or other (n=100).

Reliability and validity study data analysis. A number of measures were used to determine if the final instrument produced reliable, valid, and generalizable scores in this sample of students. An exploratory factor analysis using Varimax rotation was used to assess the validity of the instrument. Factor analysis was used to ascertain the underlying structure of the 59 behaviors (weighted by severity level), reflecting both level of severity

and type of behavior. The mild items were weighted by multiplying the item score by one, while scores for moderate items were multiplied by two and severe items by three.

Internal consistency of the items on the instrument was assessed using Cronbach's alpha (1951) and McDonald's (1985) omega calculations. McDonald's omega is a reliability coefficient that is created from the factor model for homogeneous items. It is the ratio of the common variance to the total variance (common variance plus the error variance). Omega represents the accuracy of the instrument in measuring an attribute, thus it is a measure of the reliability of an instrument.

Reliability and validity coefficients were calculated for each of the six scales identified in the factor analysis and for the overall instrument. The omega coefficient also serves as measure of the validity and generalizability of the scores for the instrument. Since scores on an instrument are valid if they measure the attribute of the participants that the instrument was designed to measure in the appropriate population, omega is also a validity coefficient. It is the square of the correlation between the score on the instrument and the score of the domain, thus, the instrument score is valid to the extent that it is related to the score for the domain defined by the attribute (McDonald, 1999). In addition, omega also is a measure of the generalizability of the instrument. It is the squared correlation between the total score (or mean) of items on an instrument and the mean score of items on a test of infinite length (McDonald, 1985, 1999). Frequencies and descriptive statistics also were calculated for each of the items on the instrument.

Results

Based on the findings from the series of pilot studies, feedback from students, and the content-related evidence, the final version of the student management instrument was created and used in the reliability and validity study. The final expert instrument used to assess the content-related evidence includes all of the items from the final version of the student management instrument and included input on the organization and classification of items from experts who participated in the expert pilot studies.

Phase Two: Content-Related Evidence

The majority of the experts' ratings were consistent with the pre-specified classifications (83%). Differences may be attributed to the situation-specific nature of some of the items, as "talking". The overall intraclass correlation coefficient (interrater reliability) was .99, suggesting strong agreement among the experts on the items composing the instrument. Expert item agreement information is available in Table 2. Experts did not indicate that there were any student troublesome behaviors missing from the survey or that any of the items were problematic.

Phase Three: Reliability and Validity

The exploratory factor analysis identified six factors that explained 50.8% of the total variance. Using only items with the highest loadings on the factor, six summative scales were constructed. These have been named as follows: (1) aggressive (e.g., bullying); (2) low engagement/irresponsibility (e.g., forgets gym clothes); (3) fails to follow directions (e.g., doesn't line up right); (4) illegal/harmful (e.g., drug use); (5)

distract/disturb/annoy others (e.g., giggling); and (6) poor self-management (e.g., late assignments). The structure coefficients for the items ranged from .26 to .73 (M=.53) see Table 3 for a complete listing of the structure coefficients. All but one of the structure coefficients exceeded 0.30, which is considered the minimum value for an item to contribute to a factor (McDonald, 1999; Safrit & Wood, 1989). A seventh factor was identified in the exploratory factor analysis. Since the seventh factor only included two items with no logical relationship to each other and due to the fact that the items each had a similar loading on another more logical scale, the items were added to the alternative scales (i.e., item 42 to scale 2 and item 34 to scale 6). These analyses support the instruments' ability to yield valid scores. The management items clustered into six logical scales.

Reliability assessments showed a high level of interitem agreement. The Cronbach's alpha reliability coefficients for scores on the six scales were .91, .88, .89, .74, .72, and .53 for scales 1-6 respectively, whereas the omega values ranged from .88 to .52, both with only one coefficient below .72. These results support the internal consistency of the scores on the instrument's scales. All of the scales met the minimum guidelines for internal consistency for research purposes (Nunnally, 1967). The scale with the lowest level of internal consistency (i.e., scale 6) has fewer items and a large conceptual range of items. The instrument's overall internal reliability coefficients of .97 (Cronbach's alpha) and .96 (McDonald's omega) highlight the strong interitem agreement among the items on the instrument. The omega results also support the validity and generalizability of scores on the scales that comprise the instrument.

Items do reflect reliable behavioral clusters reflective of severity. The first factor--offends others--includes mostly severe and moderate items. The second factor--not participating as they should--includes almost all mild items. The third factor--not following directions--includes mostly mild and moderate behaviors. The fourth factor--illegal or harmful--consists of mostly moderate and severe items. The final two factors--bothering others and extra work for the teacher--both contain mostly mild items.

Discussion

An instrument that produces reliable and valid scores has been developed that can be used in future studies to investigate students' views of classroom management in physical education. The strong agreement among the experts on the severity of the items composing the instrument and the grouping of items into logical categories provide validity evidence for the instrument. Factor analysis results suggest that the items on the instrument create six scales: (1) aggressive, (2) low engagement/irresponsibility, (3) fails to follow directions, (4) illegal/harmful, (5) distract/disturb/annoy others, and (6) poor self-management. The items on each scale also are conceptually meaningful. Inter-item agreement from both alpha and omega coefficients showed that the instrument produced reliable information. Omega coefficients also support the validity and generalizability of scores on the scales that comprise the instrument. In addition, items reflect reliable behavioral clusters of severity. The scales are comprised of items with similar severity levels.

One limitation to this study is that the validity evidence was collected based on students' self-reported data. While the results of the data analyses provide strong supporting evidence for the validity of the instrument, future studies should correlate the survey information with student behavioral information gained through observations.

The major advantage of developing and using a survey to assess student behavior is the ability to assess student misbehaviors across many teachers and contexts in a relative short period of time using an instrument that produces reliable and valid information.

There are also some disadvantages to using this type of instrument including that some detail may be missed related to actual student behaviors and that the results are based on self-report data. The optimal situation may be to use a survey instrument along with observations/interviews in order to obtain the most information about students' misbehaviors in physical education classes.

The classroom management instrument developed in this study can now be used to further investigate students' views of classroom management to begin to understand how teachers and students view classroom management in physical education. The instrument can be used to study the relationship between what students report happening in classes and actual classroom management events. It also can be used to document how students' disruptive behaviors change through intervention efforts. Information gained from using the classroom management instrument can be used to inform teacher education programs and teacher inservice programs. This will contribute to educating teachers about current student misbehaviors as well as to help them develop strategies to minimize disruptive behavior.

Future research efforts need to include an experimental assessment of the validity evidence of the instrument. To compare students' reported behaviors occurring in classes with actual behaviors. This study also should be repeated with another sample of students including a confirmatory factor analysis of the six scales. Furthermore, qualitative measures (e.g., interviews) of students' perceptions of classroom management need to be investigated. Continuing to investigate students' views of classroom management will lead to a better understanding of management issues and behaviors that interfere with student learning.

References

Borg, M. G. (1998). Secondary school teachers' perception of pupils' undesirable behaviours. British Journal of Educational Psychology, 68, 67-79.

Borg, M. G., & Falzon, J. M. (1989a). Primary school teachers' perception of pupils' undesirable behaviours. Educational Studies, 15, 251-260.

Borg, M. G., & Falzon, J. M. (1989b). Maltese primary schools teachers' perception of the seriousness of stealing, cruelty, and lying in schoolchildren. Melita Theologica, XL (2), 111-120.

Borg, M. G., & Falzon, J. M. (1993). A factor analytic study of teachers' perceptions of pupils' undesirable behaviours: A rejoinder to Langfeldt. (1992). British Journal of Educational Psychology, 63, 513-518.

Brophy, J. (1983). Improving instruction: Effective classroom management. The School Administrator, 40, 33-36.

Brophy, J., & Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Clarizio, H. F. (1976). Toward positive classroom discipline (3rd ed.). New York: John Wiley and Sons.

Cothran, D. J., & Ennis, C. D. (1997). Students' and teachers' perceptions of conflict and power. Teaching and Teacher Education, 13, 541-553.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 296-334.

Cullingford, C. (1991). The inner word of school: Children's ideas about schools. London: Cassell Educational Limited.

Depaepe, J., French, R., & Lavay, B. (1985). Burnout symptoms experienced among special physical educators: A descriptive longitudinal study. Adapted Physical Activity Quarterly, 2, 189-196.

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 392-431). New York: Macmillan.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. Review of Research in Education, 5, 163-198.

Duke, D. L., & Jones, V. F. (1984). Two decades of discipline-Assessing the development of an educational specialization. Journal of Research and Development in Education, 17, 25-35.

Farrell, E., Peguero, G., Lindsey, R., & White, R. (1988). Giving voice to high school students: Pressure and boredom, ya know what I'm sayin'? American Educational Research Journal, 25, 489-502.

Fields, B. A. (1986). The nature and incidence of classroom behaviour problems and their remediation through preventative management. Behaviour Change, 3, 53-57.

Houghton, S., Wheldall, K., & Merrett, F. (1988). Classroom behavior problems which secondary school teachers say they find most troublesome. British Educational Research Journal, 14, 297-312.

Keating, X. D., Kulinna, P. H., & Silverman, S. (1999). Measuring teaching behaviors, lesson context, and physical activity in school physical education program: Comparing the SOFIT and the C-SOFIT instruments. Measurement in Physical Education and Exercise Science, 3, 207-220.

Kennedy, E. F. (1982). Discipline in the physical education setting. Physical Educator, 39, 91-94.

Kulinna, P. H., & Silverman, S. (1999). The development and validation of scores on a measure of teachers' attitudes toward teaching physical activity and fitness. Educational and Psychological Measurement, 59, 507-517.

Kounin, J. S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart and Winston.

McCormack, A. (1997). Classroom management problems, strategies and influences in physical education. European Physical Education Review, 3, 102-115.

McDonald, R. P. (1985). Factor analysis and related methods. Mahwah, NJ: Lawrence Erlbaum.

McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum.

McNamara, E. (1985). Are the techniques of behavior modification relevant to problems of concern to teachers in secondary schools? Behavioural Approaches with Children, 9, 34-45.

Nicholls, J. G. (1992). Students as educational theorists. In D. Schunk & J. Meece (Eds.), Students' perceptions in the classroom (pp. 267-286). Hillsdale, NJ: Erlbaum.

Nunnally, J. C. (1967). Psychometric Theory. New York: McGraw-Hall.

O'Sullivan, M., & Dyson, B. (1994). Rules, routines, and expectations of 11 high school physical education teachers. In M. O'Sullivan (Ed.). High school physical education teachers: Their world of work [Monograph] Journal of Teaching in Physical Education, 13, 361-374.

Safrit, M. J., & Wood, T. M. (1989). Measurement Concepts in Physical Education and Exercise Science. Champaign, IL: Human Kinetics.

Silverman, S., & Subramaniam, P. R. (1999). Student attitude toward physical education and physical activity: A review of measurement issues and outcomes. Journal of Teaching in Physical Education, 19, 97-125.

Solmon, M. A., Cothran, D. J., Stockin-Davidson, K., Ennis, C. D., Loftus, S. J., Satina, B., & Owens, L. M. (1998). Motivation to participate, work hard, and learn: Students' perceptions of the effectiveness of motivational strategies. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Stewart, M., & Destache, D. (1992). Validity of interval recording in measuring classroom climates in physical education. Journal of Teaching in Physical Education, 11, 315-323.

Vogler, E. W., & Bishop, P. (1990). Management of disruptive behavior in physical education. Physical Educator, 47, 16-26.

Wheldall, K., & Merrett, F. (1988). Which classroom behaviours do primary school teachers say they find most troublesome? Educational Review, 40, 13-27.

Wolfgang, C. H., & Glickman. C. D. (1980). Solving discipline problems. Boston: Allyn and Bacon.

Wragg, C. M. (1995, April). Classroom management the perspectives of teachers, pupils and researcher. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Zeidner, M. (1988). The relative severity of common classroom management strategies: The student's perspective. British Journal of Educational Psychology, 58, 69-77.

Ziv, A. (1970). Children's behaviour problems as viewed by teachers, psychologists and children. Child Development, 41, 871-879.

Table 1

MILD DISTURBANCE If a student does not pay attention, listen or follow instructions; fails to dress properly for class, is late, does not participate or when participating intentionally moves slowly, shows off or only works with friends, violates minor procedural rules, or has a personal/psychological disposition that interferes with learning, please rate the behavior a mild class disturbance.
MODERATE DISTURBANCE If a student verbally interacts with others either in a friendly or unfriendly manner including inappropriate language or gestures, loses his/her temper; interrupts or interferes with the teacher or other students' work; misuses or steals personal items and equipment, cheats at games or other unsportspersonlike (unsportsmanlike) behavior, or leaves their assigned area; these behaviors can be considered a moderate class disturbance.
SEVERE DISTURBANCE If a student exhibits any action that intimidates or threatens another person; engages in unsafe actions for themselves or others; or smokes; these behaviors can be considered a severe class disturbance.

Table 2

Expert Item Agreement by Severity

Item #	Item Description	Percent of Experts Assigning Item to Severity Category			Deviation from Standard
		Mild	Moderate	Severe	
4	Gum chewing	100.0%			.00
15	Doesn't line up right	100.0%			.00
29	Acts shy and withdrawn	96.3%	3.7%		.04
42	Using menstrual period to not participate	96.3%	3.7%		.04
23	Doesn't participate because of pregnancy	96.2%	3.8%		.04
48	Lazy	92.3%	7.7%		.07
57	Dirty gym clothes	92.3%	7.7%		.07
13	Forgets gym clothes	92.3%	7.7%		.08
17	Doesn't pay attention	92.3%	7.7%		.08
19	Not following directions	92.3%	7.7%		.08
25	Poor self image	91.7%	8.3%		.08
11	Giggling	88.5%	11.5%		.12
3	Misses/late for class	88.0%	12.0%		.12
28	Forms cliques	88.9%	7.4%	3.7%	.15
33	Doesn't participate	88.9%	7.4%	3.7%	.15
34	Late assignments	85.2%	14.8%		.15
39	Continually saying "I can't do it"	85.2%	14.8%		.15
56	Moves slowly on purpose	84.6%	15.4%		.15
49	Attention seeking	84.0%	16.0%		.15
27	Pretends to be sick	81.5%	18.5%		.19
24	Showing off	80.8%	19.2%		.19
38	Whining	80.8%	19.2%		.19
58	Tattling	76.0%	24.0%		.23*
30	Always asks to have instructions repeated	74.1%	25.9%		.26*
31	Clings to the teacher	73.1%	26.9%		.27*
6	Can't sit still (hyperactive)	69.2%	30.8%		.31*
41	Always must be first or best	56.0%	40.0%	4.0%	.48*
10	Talking	65.4%	34.6%		-.65*
14	Interrupts	30.8%	69.2%		-.31*
18	Doesn't take care of equipment	23.1%	76.9%		-.23*
32	Gets upset when loses/fails	11.1%	85.2%	3.7%	-.07
35	Leaving the group during activity	11.1%	85.2%	3.7%	-.07

Table 2 Continued

1	Swearing/cursing	4.2%	91.7%	4.2%	.00
43	Poor sportsmanship		100.0%		.00
53	Sneaks out of class	7.7%	84.6%	7.7%	.00
46	Quarrelsome	74.1%	11.1%	14.8%	.04
51	"Smart mouth" toward students	7.7%	80.8%	11.5%	.04
55	Talking back	7.7%	80.8%	11.5%	.04
59	"Smart mouth" toward teacher		96.2%	3.8%	.04
9	Arguing		92.3%	7.7%	.08
26	Makes fun of other students	7.4%	74.1%	18.5%	.11
54	Keeps others from working	3.8%	80.8%	15.4%	.11
20	Lying	7.7%	73.1%	19.2%	.12
22	Temper tantrums	3.8%	80.8%	15.4%	.12
12	Writing on walls/lockers	7.7%	69.2%	23.1%	.15
47	Obscene gestures		80.8%	19.2%	.19
36	Cheating		74.1%	25.9%	.26*
7	Stealing		65.4%	34.6%	.35*
40	Gang symbols displayed		14.8%	85.2%	-.15
8	Bullying		11.5%	88.5%	-.12
2	Fighting	3.8%		96.2%	-.08
37	Playing too rough & risking injury		7.4%	92.6%	-.07
5	Smoking		3.8%	96.2%	-.04
16	Unsafe actions		3.8%	96.2%	-.04
44	Sexual harassment		3.7%	96.3%	-.04
50	Threatens others		3.8%	96.2%	-.04
52	Pushing or punching others		3.8%	96.2%	-.04
21	Bringing weapons to class			100.0%	.00
45	Drug use			100.0%	.00

* Significantly discrepant from gold standard at the p=.05 level.

Note. The gold standard mild items (27) are presented first followed by the moderate (21) and severe (11) items, respectively. Items within each group are listed by deviation from the gold standard from smallest to the largest.

Table 3

Item Numbers, Descriptions, and Factor Structure Coefficients

Item	Brief Item Description	Domains					
		I	II	III	IV	V	VI
	Scale I: Aggressive						
52	Pushing or punching others	.725					
50	Threatening others	.694					
51	Smart mouth toward students	.688					
47	Obscene gestures	.638					
55	Talking back	.610					
37	Playing too rough	.610					
59	Smart mouth toward teacher	.583					
46	Quarrelsome	.574					
26	Makes fun of other students	.540					
54	Keeps others from working	.507					
9	Arguing	.479					
44	Sexual harassment	.458					
36	Cheating	.458					
41	Always must be first or best	.451					
8	Bullying	.443					
2	Fighting	.349					
	Scale II: Low Engagement/Irresponsibility						
57	Dirty gym clothes		.650				
38	Whining		.620				
39	Continually saying, "I can't do it"		.577				
43	Poor sportsmanship		.538				
48	Lazy		.511				
25	Poor self-image		.507				
56	Moves slowly on purpose		.486				
13	Forgets gym clothes		.485				
49	Attention seeking		.473				
29	Acts shy and withdrawn		.456				
58	Tattling		.453				
33	Doesn't participate		.449				
24	Showing off		.416				
27	Pretends to be sick		.364				
42	Using menstrual period as an excuse		.358				
35	Leaving group during activity		.330				

	Scale III: Fails to Follow Directions				
17	Doesn't pay attention	.674			
19	Not following directions	.670			
18	Doesn't take care of equipment	.624			
15	Doesn't line up right	.589			
16	Unsafe actions	.558			
14	Interrupts	.551			
20	Lying	.415			
30	Always asks to have instructions repeated	.397			
	Scale IV: Illegal/Harmful				
45	Drug use		.681		
5	Smoking		.673		
21	Bringing weapons to class		.667		
40	Displaying gang symbols		.602		
12	Writing on walls		.552		
23	Not participating due to pregnancy		.546		
7	Stealing		.506		
53	Sneaking out of class		.470		
	Scale V: Distract/Disturb/Annoy Others				
10	Talking			.717	
11	Giggling			.696	
4	Gum chewing			.582	
1	Swearing			.484	
6	Can't sit still (hyperactive)			.465	
28	Forming cliques			.405	
3	Misses or is late for class			.340	
	Scale VI: ***Poor Self-Management***				
31	Clings to the teacher				.556
32	Gets upset when loses/fails				.533
22	Temper tantrums				.401
34	Late assignments				.257

Note. n=2309